



2110412

## Parallel Computer Architecture: Multi-core Tech.

3: Concepts from Jouppi's paper (ILP)

Krerk Piromsopa, Ph.D.

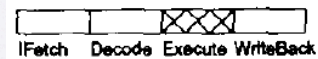
Department of Computer Engineering  
Chulalongkorn University

## Definitions

- Operation latency (OL):  
the number of cycles until the result of an instruction is available.
- (Instruction) Issue Latency (IL):  
the number of cycles required between issuing two consecutive instructions.
- (Instruction) Issue Rate (Issue Parallelism) (IP):  
the maximum number of instructions that can be issued in a cycle. (degree)
- Maximum Instruction-level Parallelism (MILP):  
the maximum number of simultaneously executing instructions

## Notion

Key:



- Four-stage instruction pipeline (K)
  - IF - Instruction Fetch
  - DE - Instruction Decode
  - EX - Execute
  - WB - Write Back

## Base Scalar Machine

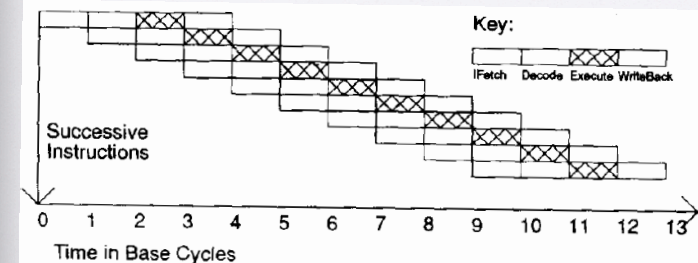


Fig. 1. Execution in a base machine.

- OL = 1 cycle
  - IL = 1 cycle
  - IP = 1 instruction / cycle
  - MILP = K
- Time required to execute N instruction?
- $$T(IP, OL) = T(1, 1) = K + (N-1)$$

# SuperScalar Machine

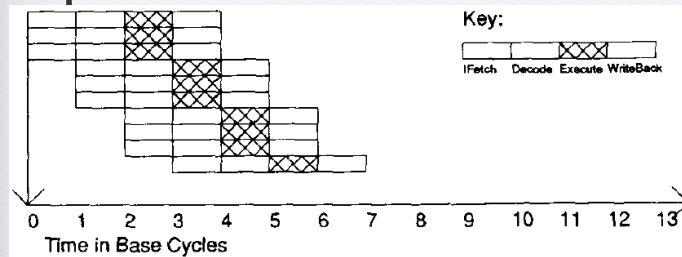


Fig. 4. Execution in a superscalar machine ( $n = 3$ ).

- OL = 1 cycle
  - IL = 1 cycle
  - IP =  $n$  instruction / cycle
  - MILP =  $nK$
- Time required to execute  $N$  instruction?
- $$T(IP, OL) = T(n, 1) = K + (N-n)/n$$
- Speedup?

- Speedup:

$$\begin{aligned} S(n, 1) &= T(1, 1) / T(n, 1) \\ &= (K + (N-1)) / (K + (N-n)/n) \\ &= n(N+K-1) / N + n(K-1) \end{aligned}$$

- More  $N$ ,  $S(n, 1) \rightarrow n$

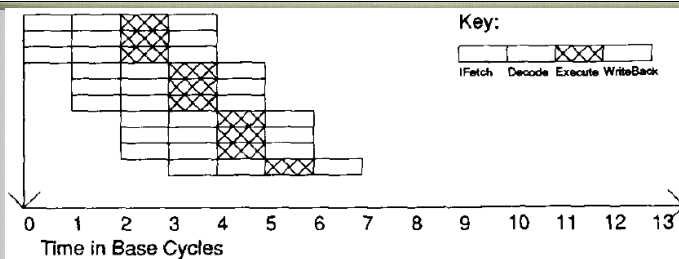


Fig. 4. Execution in a superscalar machine ( $n = 3$ ).

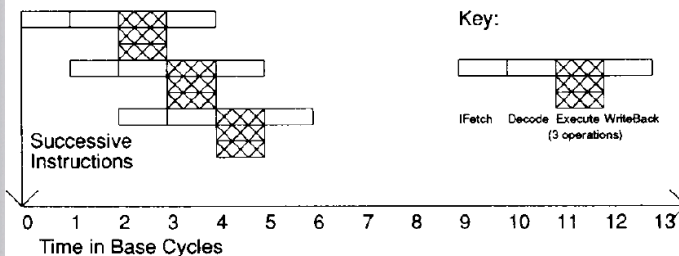


Fig. 5. Execution in a VLIW machine.

# SuperPipelined Machine

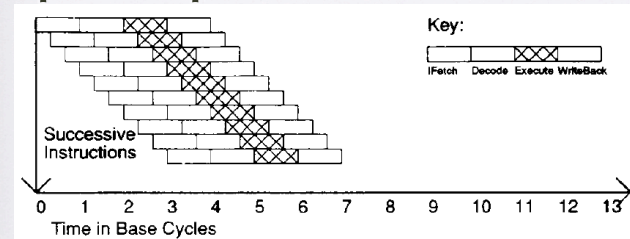


Fig. 6. Superpipelined execution ( $m = 3$ ).

- OL =  $m$  minor cycles  
= 1 baseline cycle
  - IL = 1 minor cycle  
=  $1/m$  baseline cycle
  - IP = 1 instruction / minor cycle  
=  $m$  instructions / baseline cycle
  - MILP =  $mK$
- Time required to execute  $N$  instruction?
- $$T(IP, OL) = T(1, m) = K + (N-1)/m$$
- Speedup?



- Speedup:

- $S(1,m)$   
 $= T(1,1) / T(1,m)$   
 $= (K+(N-1)) / (K+(N-1)/m)$   
 $= m(N+K-1) / (mK + N - 1)$

- More N,  $S(1,m) \rightarrow m$

## SuperPipelined SuperScalar (SS) Machine

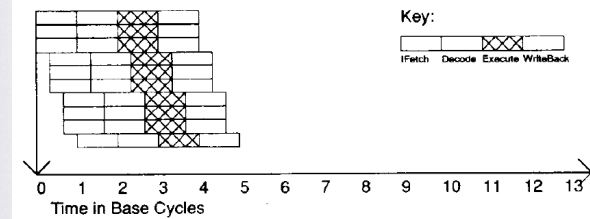


Fig. 7. A superpipelined superscalar ( $n = 3, m = 3$ ).

- $OL = m$  minor cycles  
 $= 1$  baseline cycle
  - $IL = 1$  minor cycle  
 $= 1/m$  baseline cycle
  - $IP = n$  instruction / minor cycle  
 $= nm$  instructions / baseline cycle
  - $MILP = nmK$
- Time required to execute N instruction?  
 $T(IP, OL) = T(n,m) = K + (N-n)/nm$
- Speedup?

- Speedup:

- $S(n,m)$   
 $= T(1,1) / T(n,m)$   
 $= (K+(N-1)) / (K+(N-m)/nm)$   
 $= nm(N+K-1) / (nmK + N - 1)$

- More N,  $S(n,m) \rightarrow nm$

## Vector Machine

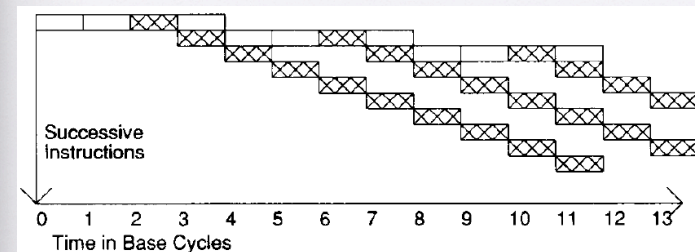


Fig. 8. Execution in a vector machine.

# Supersymmetry

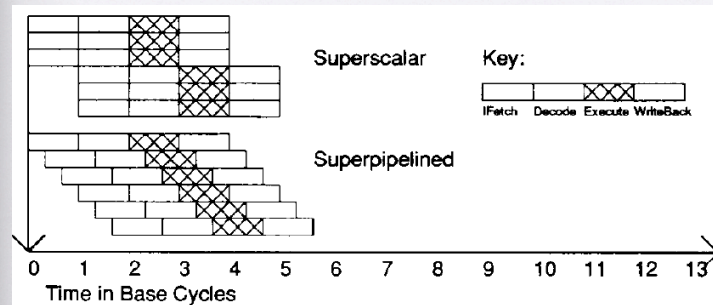
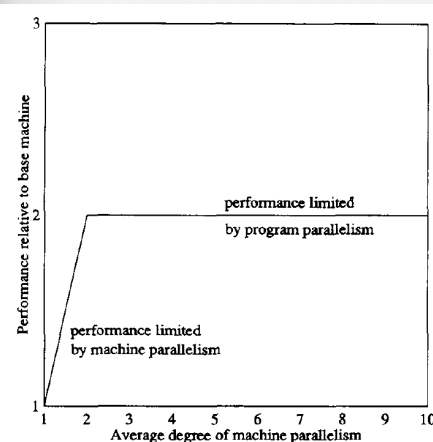


Fig. 10. Startup in superscalar versus superpipelined.

## Limitations

- Limitations in ILP affect Issue Parallelism (Benchmark Parallelism) - BP
- Limitations in available hardware resources affect MP (Machine Parallelism)
- $MP = Sp \times Ss$   
 $Sp$  - Average degree of superpipelining  
 $Ss$  - Degree of parallel issue (IP)

## Limitations



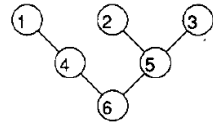
$$\text{Performance} = \min(\text{MP}, \text{BP})$$

## Variations in Operation Latency (Limit on BP)

Cycle	Uniform	Serial Non-uniform	Parallel Non-uniform
	All operations have latency=2	Odd operations have latency=3, even operations have latency=1	Operations $\leq 4$ have latency = 3, operations $> 4$ have latency = 1
1	1,2	1,2	1,2
2		4	
3	3,4	6	
4		3,8	3,4
5	5,6		
6		5	5,6
7	7,8		7,8
8	— execution complete		— execution complete
9		7	
10			
11			
12			
13		— execution complete	



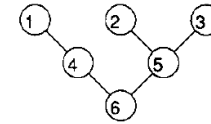
## Variations in # ILP (Limit on MP)



Cycle	Instructions issued	
	Unlimited superscalar	Degree 2 superscalar
1	1,2,3	1,2
2	4,5	3,4
3	6	5
4		6

Fig. 18. Variations in the number of instructions executable in parallel.

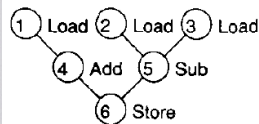
## Variations in Instruction Class (Limit on MP)



Cycle	Instructions issued	
	Unlimited superscalar	Degree 2 superscalar
1	1,2,3	1,2
2	4,5	3,4
3	6	5
4		6

Fig. 18. Variations in the number of instructions executable in parallel.

## Variations in Instruction Class (Limit on MP)



Cycle	Instructions issued	
	Classless superscalar	Mem ALU branch class superscalar
1	1,2,3	1
2	4,5	2,4
3	6	3
4		5
5		6

Fig. 23. Nonuniform distribution by instruction class.

The End  
Questions?